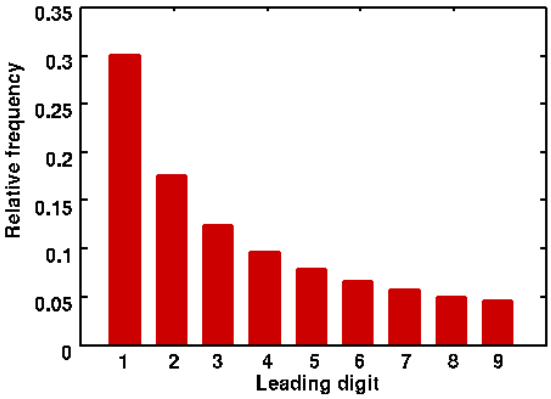
Benford's Law



One remarkable feature of numbers used in the real world (measured in physical units like dollars, meters, grams etc) is that the first leading digit of a collection of these numbers is not random, although the numbers, taken together, may seem so.

For example, in the sequence $123.43, $327.30, $457.67, $1327.03 the first digits are 1, 3, 4 and 1. We could plot this in a histogram so that '1' has a frequency of 2, '3' has a frequency of 1 and '4' has a frequency of 1. When tallied for an entire tax return with several hundred numbers, the frequency table follows Benford's Law.

The Internal Revenue Service uses Benford's Law to identify tax cheats! For example, if we take all the numbers (in dollar units) you enter in a tax form and count up the frequency of 1's ,2's ,3's …9's, we will get a histogram that shows that 1's appear 30% of the time, 2s appear 17% of the time, 3s appear 12% of the time, and so on.

American astronomer Simon Newcomb noticed that in logarithm books (used at that time to perform calculations), the earlier pages (which contained numbers that started with 1) were much more worn than the other pages.

What is interesting about this law is that, for example, if you made one measurement in inches and computed the first-digit frequency, you would get the same first-digit histogram if you then changed the units to meters or any other unit of measurement!

Does the universe obey this law too?

Below are some archives of data online. Tally the first digit frequencies and test if Benford's Law holds.

Problem 1: **Sunspot Numbers** (Pure numbers: no units) from the National Geophysical Data Center:
http://www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html#american

Problem 2: **Solar Wind Magnetism**: NASA ACE satellite measurements of the solar wind.
http://www.swpc.noaa.gov/ftpmenu/lists/ace2.html
Open a file with 'mag' as part of the file name such as 200712_ace_mag_1h.txt
Look at Column 11 labeled 'Bt' which are the measurements of the total solar wind magnetic field strength in units of nanoTeslas.

Problem 3: **Depth of latest earthquakes** in kilometers from the USGS:
http://earthquake.usgs.gov/eqcenter/recenteqsww/Quakes/quakes_all.html

Inquiry Questions: Can you find any patterns in the frequency histograms for the above problems? Why don't the digits appear with about the same frequency? Can you think of other astronomical data that you can use to further test your hunches, or refine any patterns you may have found?

Answer Key:

Problem 1: **Sunspot Numbers** (Pure numbers: no units) from the National Geophysical Data Center:
http://www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html#american

Answer:
Here are some sample numbers for the month of April 1998

53, 55, 48, 47, 57, 61, 91, 97, 103, 102, 93, 75, 61, 60, 63, 56, 34, 26, 30, 31, 31, 24, 17, 19, 14, 13, 13, 27, 30, 39

The leading digits are 5,5,4,4,5,6,9,9,1,1,9,7,6,6,6,5,3,2,3,3,3,2,1,1,1,1,1,2,3,3
Ordered from 1 to 9 by frequency: 7,3,6,2,4,4,1,0,3
Note: There are significantly more 1s.

Problem 2: **Solar Wind Magnetism**: NASA ACE satellite measurements of the solar wind.

http://www.swpc.noaa.gov/ftpmenu/lists/ace2.html
Open a file with 'mag' as part of the file name such as 200712_ace_mag_1h.txt
Look at Column 11 labeled 'Bt' which are the measurements of the total solar wind magnetic field strength in units of nanoTeslas.

Answer:
Sample numbers from the above file are: 4.7, 5.4, 4.7, 1.6, 1.6, 2.2, 2.7, 1.9, 2.5, 3.9, 3.7, 3.6, 3.7, 3.0, 3.1, 2.4, 2.7, 2.6, 1.2, 3.1, 4.2, 3.6, 3.6, 3.6, 1.1, 2.8, 2.1, 3.4, 3.4, 2.4, 2.2, 1.7, 0.4, 2.7, 2.8

The leading digits are: 4,5,4,1,1,2,2,1,2,3,3,3,3,3,3,2,2,2,1,3,1,2,3,3,3,1,2,2,3,3,2,2,1,4,2,2
Ordered from 1 to 9 by frequency: 7,13,12,3,1,0,0,0,0
Note: this data is bounded by a maximum value near 6, so it does not show Benford's Law.

Problem 3: **Depth of latest earthquakes** in kilometers from the USGS:

http://earthquake.usgs.gov/eqcenter/recenteqsww/Quakes/quakes_all.html

Answer:
Sample depths: 10.0, 36.5, 12.5, 64.6, 3.9, 10.2, 10.0, 10.0, 98.4, 3.5, 500.7, 92.7, 81.6, 22.4, 2.0, 46.0, 53.3, 100.0, 10.0, 49.1, 57.0

The leading digits are: 1,3,1,6,3,1,1,1,9,3,5,9,8,2,2,4,5,1,1,4,5
Ordered from 1 to 9 by frequency: 7,2,3,2,3,1,0,1,2
Note: this data does show Benford's Law because it is not apparently bounded.

Inquiry Questions: Students should discover that numbers that are limited in some way, such as the height of your students (all about the same height) will not show Benford's law. Only numbers that are not in some way bounded above or below in value (such as solar wind magnetism, which has a rather narrow range of values near 5.0 nTeslas). Also, street address and other pure numbers do not follow Benford's Law, only numbers related to unbounded, measured parameters with physical units (temperature ,density, mass ,energy, length, time, etc) will work. Can the student's discover why this is so?